Please give well-motivated answers.

1 (2 points). Prove that for any source $X \sim p(x)$ with $x \in \mathcal{X}$ and any binary prefix-free code with lengths $l(x), x \in \mathcal{X}$, we have the relationship

$$E[l(X)] = H(X) + D(p(x)||r(x)) - \log c,$$

where $r(x) = 2^{-l(x)}/c$ and $c = \sum_{x \in \mathcal{X}} 2^{-l(x)}$. Conclude that, if the distribution is dyadic (i.e., if $p(x) = 2^{-k(x)}$ for integers $k(x)$), then we can find a prefix-free code with average length equal to $H(X)$ (Hint: Write explicitly the right-hand side of the equality above).

**Sol.:**

$$- \sum_{x \in \mathcal{X}} p(x) \log p(x) + \sum_{x \in \mathcal{X}} p(x)(\log (p(x)) - \log(r(x))) = E[l(X)].$$

If the distribution is dyadic, then we can choose $l(x) = k(x)$ (since this satisfies Kraft’s inequality), and thus $r(x) = 2^{-k(x)} = p(x)$ since $c = 1$. It follows that $D(p(x)||r(x)) = 0$ and thus $E[l(X)] = H(X)$.

2. (1 point) Give an example of a source for which the rate required by a Shannon code is close to $H(X) + 1$.

**Sol.:** Consider a source $X \sim Ber(\epsilon)$ for a very small $\epsilon > 0$. Then, we have $H(X) \simeq 0$, and $R = 1$ for a Shannon code.

3 (3 points). Consider a source $X^n \sim p(x^n)$ ($x^n \in \mathcal{X}^n$), and any fixed-to-fixed source code with rate $R$ (i.e., encoder $W(X^n)$ and decoder $\hat{X}^n(W)$ with $W$ consisting of $nR$ bits). The probability of error is $P_e = \Pr[\hat{X}^n \neq X^n]$. We want to prove the inequality

$$P_e \geq \Pr \left[ -\frac{1}{n} \log p(X^n) \geq R + \gamma \right] - 2^{-n\gamma} \tag{1}$$

for any such code and any $\gamma > 0$. To this end, define $B$ as the set of sequences $x^n$ that the code reproduces correctly and $T$ as the set $\{x^n : -\frac{1}{n} \log p(x^n) \geq R + \gamma\}$, and answer the following.

a. Show that $\Pr[T] \leq P_e + \Pr[T \cap B]$ (Hint: The events $B$ and $B^c$ form a partition of the probability space).

b. Prove the upper bound $\Pr[T \cap B] \leq 2^{-n\gamma}$ (Hint: Use the definition of $T$ and the cardinality of $B$).

c. Point a. and b. prove (1). Now, use (1) to show that if $R < H(X)$, then $P_e \rightarrow 1$ for $n \rightarrow \infty$. 

\textbf{Soli.}: a. We have
\[
\Pr[T] = \Pr[T \cap B^c] + \Pr[T \cap B] \\
\leq \Pr[B^c] + \Pr[T \cap B] \\
= P_e + \Pr[T \cap B],
\]
where we have used the fact that $P_e = \Pr[B^c]$.
b. We have
\[
\Pr[T \cap B] \leq |B|2^{-n(R+\gamma)} \\
\leq 2^{nR}2^{-n(R+\gamma)} \\
= 2^{-n\gamma},
\]
where the first inequality follows by the definition of $T$ (every sequence in $T$ satisfies $p(x^n) \leq 2^{-n(R+\gamma)}$). The second inequality follows since $|B| \leq 2^n R$.
c. For sufficiently small $\gamma$, if $R < H(X)$, by the law of large numbers, we have that
\[
\Pr[\frac{1}{n}\log p(X^n) \geq R + \gamma] \to 1.
\]

\textbf{4.} (2 points) a. Calculate the entropy rate of the process $X_k = X_{k-1} \oplus Z_k$ with $Z_k \sim \text{Ber}(p)$ and i.i.d. ($X_k$ is assumed to be stationary).
b. Repeat for $X_k = X_{k-1} \oplus X_{k-2} \oplus Z_k$.

\textbf{Soli.}: a. Since $X_k$ is stationary, we can write
\[
H(\mathcal{X}) = \lim_{n \to \infty} H(X_n|X_1, \ldots, X_{n-1}) \\
= H(X_n|X_{n-1}) \\
= H(Z) \\
= H(p).
\]
b. Similarly, we have
\[
H(\mathcal{X}) = \lim_{n \to \infty} H(X_n|X_1, \ldots, X_{n-1}) \\
= H(X_n|X_{n-1}, X_{n-2}) \\
= H(Z) \\
= H(p),
\]
where the second equality follows since, given $X_{n-1}, X_{n-2}, X_n$ does not depend on the samples $X_{n-k}$ with $k > 3$.

\textbf{5.} (2 points) Find a Huffman code and a Shannon code for the source $(1/3, 1/5, 1/5, 2/15, 2/15)$. Compare their average length.

\textbf{6.} (2 points) Consider random variable $X \sim \text{Ber}(0.5)$, and a random variable $Y$ distributed as follows: if $X = 0$, $Y$ equals 0 with probability 0.7 and 1 with probability 0.3; if $X = 1$, $Y$ equals 1 with probability 0.7 and 0 with probability 0.3.
a. Find the function $\hat{X} = f(Y) \in \{0, 1\}$ that minimizes $\Pr[\hat{X} \neq X]$. 

b. For the given estimator \( \hat{X} = f(Y) \), calculate \( H(X|\hat{X}) \) and \( P_e \).

c. Compare the results at the previous point with Fano inequality.

\textit{Sol:} a. By inspection of the joint distribution, we have

\[
\begin{align*}
 f(0) &= 0 \\
 f(1) &= 1.
\end{align*}
\]

b. We have

\[
H(X|\hat{X}) = \Pr[\hat{X} = 0]H(X|\hat{X} = 0) + \Pr[\hat{X} = 1]H(X|\hat{X} = 1)
= \Pr[Y = 0]H(X|Y = 0) + \Pr[Y = 1]H(X|Y = 1)
= H(0.7) = 0.8813.
\]

\[
P_e = \Pr[X = 0, Y = 1] + \Pr[X = 1, Y = 0] = 0.3.
\]

c. The Fano inequality is

\[
H(X|\hat{X}) = 0.8813 \leq H(P_e) + P_e \log(2 - 1)
= H(0.3) = 0.8813.
\]

which is thus satisfied with equality.

\textbf{7.} (2 points) Consider an i.i.d. source \( X^n \) with a \( Ber(0.3) \) distribution.

a. If \( k \) is the number of ones in a sequence \( x^n \) with \( n = 5 \), for which values of \( k \) we have that \( x^n \in A_{\epsilon}^{(5)}(X) \) for \( \epsilon = 0.2 \)?
b. Characterize the smallest set \( B \) of sequences with probability at least 0.24.
c. How many sequences are in the intersection between \( A_{\epsilon}^{(5)}(X) \) for \( \epsilon = 0.2 \) and \( B \)?

\textit{Sol:} We have \( H(X) = 0.8813 \) bits. By definition of typical set, we need to verify that

\[
0.6813 \leq -\frac{1}{5} \sum_{i=1}^{5} \log p(x_i) \leq 1.0813.
\]

We can calculate the following probabilities:

- for sequences with \( k = 0 \), \(-\frac{5}{5}(\log_2 0.7) = 0.5146\);
- for sequences with \( k = 1 \), \(-\frac{1}{5}(4 \log_2 0.7 + \log_2 0.3) = 0.7591\);
- for sequences with \( k = 2 \), \(-\frac{1}{5}(3 \log_2 0.7 + 2 \log_2 0.3) = 1.0035\);
- for sequences with \( k = 3 \), \(-\frac{1}{5}(2 \log_2 0.7 + 3 \log_2 0.3) = 1.2480\);
- for sequences with \( k = 4 \), \(-\frac{1}{5}(1 \log_2 0.7 + 4 \log_2 0.3) = 1.4925\);
• for sequences with $k = 5$, $-\frac{5}{5}(\log_2 0.3) = 1.7370$.

Therefore, the set $A_{0.1}^{(5)}(X)$ contains all sequences with $k = 1$ and $k = 2$ ones.

b. This contains the sequence with $k = 0$ and one of the sequences with $k = 1$. The probability of this set is $0.7^5 + 0.7^4 \cdot 0.3 = 0.2401$. Note that the most likely sequence alone ($k = 0$) has probability $0.7^5 = 0.168$.

c. Only one sequence, namely one with $k = 1$. 